DOCUMENT RESUME

ED 220 529 TM 820 551

AUTHOR Gabriel, Roy M.
TITLE Academic Growth Expectations under Various Testing Options in Title I: How Do We Interpret Aggregate Results?
PUB DATE Mar 82
NOTE 55p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Achievement Gains; Compensatory Education; Elementary Secondary Education; *Expectation; Norm Referenced Tests; Pretests Posttests; *Program Evaluation; School Districts; Scores; Student Development; *Test Interpretation; Test Norms; *Test Use
IDENTIFIERS *Expected Growth; *Normal Curve Equivalent Scores; Title I Evaluation and Reporting System

ABSTRACT

A method for translating the achievement gains in Normal Curve Equivalents (NCE) to a more conceptually understandable index of academic growth is described. This method calculates "Expected Growth (EG)" and expresses the relationship of observed achievement gains. The method is shown to be easily applied and well suited to local interpretation. Some issues when used on a more general level across several projects, districts or states are discussed. Expected Growth values are calculated using pre/posttest results and increasing the raw or scaled score (ESS) as needed to maintain the same percentile, or NCE. The increase as a proportion of the pretest standard deviation can be converted to any other standard score metric when multiplied by the standard deviation of the desired standard score scale. Expected Growth values are calculated for a variety of tests and conditions; interpretations of differences are discussed, and the implications of these differences for both large scale evaluation reporting and local use are shown. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220529

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ACADEMIC GROWTH EXPECTATIONS UNDER VARIOUS TESTING OPTIONS IN TITLE I:

HOW DO WE INTERPRET AGGREGATE RESULTS?

Roy M. Gabriel

Title I Evaluation Technical Assistance Centers
Northwest Regional Educational Laboratory

Paper presented at the annual meeting of the American Educational Research Association, March 19-23, New York, New York.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.M. Gabriel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 551

ACKNOWLEDGMENTS

The work represented in this and other related papers by the author has involved seemingly endless searches and near seizures from an array of test publishers' norms tables and technical manuals. Many colleagues have shared in this eyestrain-inducing endeavor, and I would like to publicly thank them for their efforts. To Jose Abalos, Eric Jones, Ann Davis and Lynde Paule goes my appreciation, and pledge to continue striving for the simplicity contained in the conclusion of this article--a simplicity which will help ensure they need never again repeat their painstaking efforts. Many others have offered helpful reactions and insights at earlier stages, and I am especially grateful to Dr. Rick Johnson who consistently helped me see the forest through the many trees of computer output; and to Norm Constantine, who helped generate the trees. Finally, the patience and perfectionism of Patti Wallin and Tamara West-Smith in preparing the tables and graphics has been of great and recurring value in presenting these results in this paper and in other media to a variety of audiences.

In assessing student growth using a standardized norm-referenced test (NRT), there are a variety of score metrics for expressing the achievement gains reflected by that instrument. While a range of alternatives may be beneficial in some respects, it can also lend confusion to the resolution of what otherwise seems like a very straightforward question--how much have the students learned this year? The use of norm-referenced score metrics such as percentiles or normal curve equivalents (NCEs) in reporting these results places them in a nationwide perspective. They indicate how much the students have learned in relation to how much their grade level peers in the test's norming sample have learned. Interest at the local level, however, often requires the translation of these results to other metrics--raw score gains, months growth as measured by grade equivalents, objectives mastered, etc. While these may provide a more locally interpretable reflection of student growth, the norm-referenced question inevitably rises again. That is, given that the students in question exhibited six months of academic growth during the course of the school year, how does this compare to what would be expected of them? How much do other students like these grow during a school year? The circularity of the issue is obvious.

The Title I Evaluation and Reporting System (TIERS) represents a concrete, structured approach to large-scale evaluation of programs being conducted in public and private schools across the country. Its reporting metric, the NCE, has the advantage of providing a national norm-referenced, aggregate view of the effects of Title I programs nationwide. Furthermore, the NCE scale is equal interval, making it reasonable to perform arithmetic calculations on its values. It has the

limitation described above, however--local school personnel lack a conceptual understanding of the academic growth indicated by a given achievement gain expressed in NCEs.

This paper first outlines and illustrates a method for translating the achievement gains in NCEs to a more conceptually understandable index of academic growth. The method is easy to apply and especially well suited to local interpretation. When used at a more general level (e.g., combining results across several projects, districts, or states), however, it evokes some interesting issues. Many of these are the very issues confronted in the development of TIERS in the mid-1970's.

Related Literature

With the initial development of the TIERS models and the selection of the NCE as their reporting metric (Tallmadge & Wood, 1976), a number of methodological issues arose and received considerable attention in the literature (e.g., Jaeger, 1979; Linn, 1979; Wiley, 1979). While these brought to light some limitations of the proposed evaluation system, the models and testing procedures began being implemented by many schools and districts across the country almost immediately. Full nationwide implementation was at hand by 1978. Once the evaluation procedures became more familiar and routine, issues regarding the interpretation of the results yielded by the system began to surface among practitioners. What do these NCE gains really mean? Several rules of thumb quickly emerged. They ranged from statements that indicated any NCE gain greater than zero represents positive impact of the program, and is therefore good (Tallmadge, 1978) to a standard of an NCE gain of 7, which was about one-third of a standard deviation of the NCE scale, to considerations for

the magnitude of sampling and measurement error in the estimate of gains (Davis, Deck and Demaline, 1981).

In addition to these rules of thumb, a number of empirical investigations were conducted to provide a better understanding of reasonable growth expectations of students in Title I programs. Most of these began with the expectation represented in the most commonly applied TIERS model--the Norm-Referenced model. It held that students were expected to maintain the same percentile (or NCE) at the end of the program as they had obtained before the program had begun. Tallmadge & Fagan (1977) charted the academic growth needed to meet this "equipercentile" expectation from grades 2 through 8 on seven commonly used standardized tests of reading and math. The growth in expanded scale score (ESS) units needed to maintain the same pre-post percentile was documented and expressed as a proportion of the third grade standard deviation of each test. These growth expectations were then averaged across tests to yield growth curves at each decile from grades 2-8 in reading and math. They indicated the number of standard deviation units a student needed to grow to maintain the same percentile rank year to year. The authors noted that growth expectations decreased with increasing grade levels, and were more variable at extremely high or low pretest deciles.

Stenner, Hunter, Bland & Cooper (1978) developed a similar index, termed the "standardized growth expectation" (SGE) using a slightly different approach. Rather than charting the scale score gain needed to maintain the same percentile, they documented the percentile loss incurred by maintaining the same raw score from pretest to posttest. They observed similar differences in expected growth at different grade

levels as did previous authors. In their investigation, however, they noted a major source of variation not emphasized in previous studies of growth expectation. It concerned differences in SGE among the five standardized tests they included. Reminiscent of Jaeger's (1977) concerns for aggregating evaluation results across different tests, the authors found these discrepancies "almost as striking as the grade to grade variations" (p. 7). Unfortunately, the authors considered only fiftieth percentile pretest status. While the discrepancies among tests were provocative, previous work cited here would have predicted even greater discrepancies at more extreme percentile points.

Possible uses of these expected growth indices have been suggested in the literature. Estes (1979) examplified their use as input for establishing pre-program objectives for achievement gains. For example, if program personnel aspire to increase their students' achievement rate by 25% as a result of the program, the expected growth calculation can be used in conjunction with this objective to state it in terms of NCE (or percentile) gains. Alternatively, these objectives can be based upon data other than staff aspirations. The amount of additional time or expenditures allocated to program students can be used to set a goal for additional growth of the students. For example, if compensatory education students receive six hours of reading instruction per week, as opposed to five hours for non-compensatory students, it may be reasonable to set 20% additional growth as a goal for the program. Finally, Gabriel (1981) suggested expressing an observed achievement gain as a proportion of expected growth as a more conceptually understandable measure of evaluation outcome to local audiences not familiar with the technicalities of other norm referenced metrics.

Purpose of this Paper

The purpose of the current paper is first to clearly explicate a method for calculating "expected growth" and expressing observed achievement gains in relation to it. Secondly, the expected growth values are calculated for a variety of tests and testing conditions, and interpretation of important differences is offered. Finally, the implications of these differences for both large scale evaluation reporting and local use of evaluation results are discussed.

Method

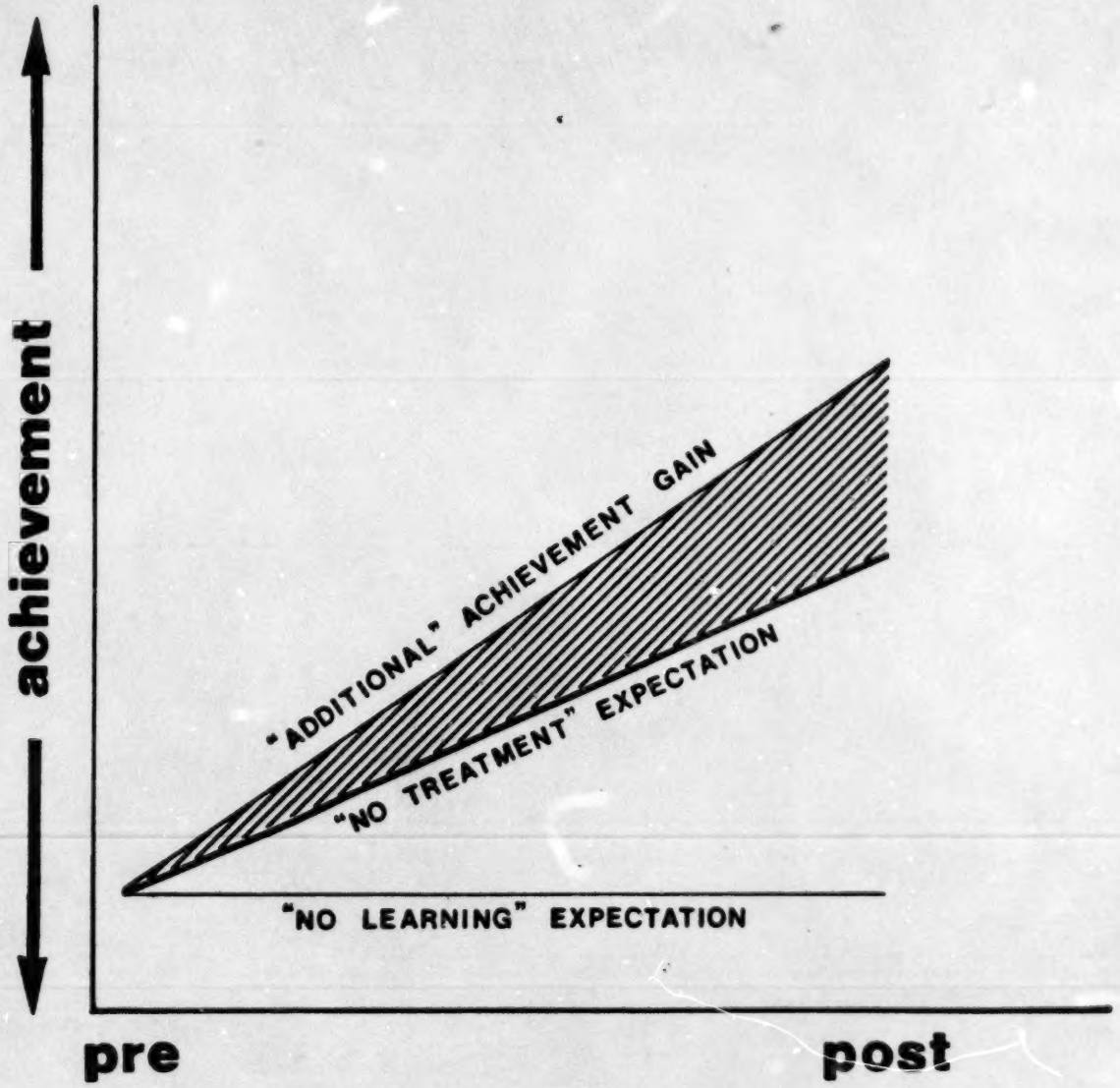
Consider the diagram of pre/post achievement test results shown in Figure 1. The horizontal line is Stenner et al's (1978) "no learning" expectation. The same raw score (or ESS) is obtained at pretest and posttest, with a consequent drop in percentile or NCE standing. The moderately sloped line represents the "no treatment expectation" (Tallmadge & Wood, 1976). The increase in raw or scaled score indicated is needed to maintain the same percentile or NCE. This increase in ESS, when expressed as a proportion of the pretest standard deviation, is the "expected growth" (EG) index used in this paper. The more steeply sloped line in Figure 1 exemplifies an "additional" achievement gain beyond what was expected. Raw or scaled scores have increased to such a degree as to evidence a percentile/NCE increase as well.

Insert Figure 1 about here

---

figure 1

ADDITIONAL GROWTH
"EXPECTED" GROWTH



The application of this method for determining a growth expectation involves three steps, referencing the norms tables of the publisher-recommended level of the particular test used. These are:

1. Determine the ESS needed at posttest to maintain the same percentile/NCE observed at pretest. Call this the "expected" posttest ESS.

2. Subtract the observed pretest ESS from the expected posttest ESS. This yields the "expected gain" in ESS.

3. Divide the expected ESS gain by the pretest ESS standard deviation to obtain the expected growth index. This provides EG as a proportion of a standard deviation (i.e., standard z scores). It can be converted to any other standard score metric by multiplying it by the standard deviation of the desired standard score scale (e.g., to NCEs by multiplying EG by 21.06).

In representing achievement growth expectations in this way, another method of expressing and interpreting gains is suggested. The additional growth represented by the achievement gain (shaded area) can be expressed as a proportion of the expected growth. The ease of interpretation is obvious. While the meaning of a 5 NCE gain may still escape school personnel or parents, 50% additional growth due to the program does not.

The appeal of such an index--percent additional growth (PAG)--is reminiscent of that of grade equivalent scores. The limitations and caveats may be just as burdensome, however. The present study will also demonstrate these.

Results

Using the method for calculating the Expected Growth index described above, values were obtained for grades two through twelve in total reading and total math as measured by five commonly used standardized tests. Pretest levels representing both the "average" student at a grade level (i.e., 50th percentile) and a typical compensatory education

student (i.e., 20th percentile) were used. Two pretest-posttest intervals were represented: fall-spring and spring-spring. The EG values were expressed in both standard deviation units (standard z scores) and NCEs (using its standard deviation of 21.06), and are presented in Tables 1-4.

Insert Tables 1-4 about here

The average EG values appearing at the far right in Tables 1-4 (i.e., averaged across tests for each combination of subject matter, grade level and percentile status) are roughly equivalent to those reported in previous studies where testing conditions represented are the same (i.e., Tallmadge & Fagan, 1977; Stenner et al., 1978). Perhaps more importantly, some general trends are replicated. A gradual decline in EG with increasing grade levels, and higher EG values for annual testing than fall-spring testing are two of the most apparent. To investigate these and other trends, the average EG values are presented graphically in Figures 2-5. In both reading and math, the decline in EG across grades is far steeper for annual (spring-spring) testing than fall-spring testing, due primarily to higher EG in the lower grade levels in the annual testing condition. The gap between them narrows in the higher grades. Contrasting the two subject matters, growth expectations are typically higher in math, though this discrepancy also disappears in the high school years. The dissipation of differences at the high school level comes as no surprise, since much of the variation in cognitive growth takes place during elementary school years.

Table 1
Expected Growth Indices
in Standard Deviation Units & NCEs
Total Reading
Fall - Spring Testing

Grade	%ile	CAT '77		CTBS '75		ITBS '78*		MAT '78		SRA '78		AVG.	
		SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE
2	20	.60	12.64	.44	9.27	.53	11.16	.73	15.37	.82	17.27	.62	13.14
	50	.60	12.64	.54	11.37	.60	12.64	.56	11.79	.54	11.37	.57	11.96
3	20	.51	10.74	.41	8.64	.41	8.64	.27	5.69	.47	9.90	.41	8.72
	50	.48	10.11	.44	9.27	.41	8.64	.23	4.84	.39	8.21	.39	8.21
4	20	.22	4.63	.20	4.21	.39	8.21	.34	7.16	.25	5.27	.28	5.90
	50	.47	9.90	.30	6.32	.39	8.21	.28	5.90	.33	6.95	.35	7.46
5	20	.19	4.00	.26	5.48	.37	7.79	.24	5.05	.11	2.32	.23	4.93
	50	.27	5.69	.30	6.32	.37	7.79	.23	4.84	.27	5.69	.29	6.07
6	20	.16	3.37	.10	2.11	.24	5.05	.02	.42	.35	7.37	.17	3.66
	50	.23	4.84	.20	4.21	.29	6.11	.19	4.00	.41	8.64	.26	5.56
7	20	.16	3.37	.13	2.74	.22	4.63	.21	4.42	.27	5.69	.20	4.17
	50	.19	4.00	.20	4.21	.26	5.48	.19	4.00	.26	5.48	.22	4.63
8	20	.15	3.16	.20	4.21	.20	4.21	.19	4.00	.29	6.11	.21	4.34
	50	.20	4.21	.17	3.58	.24	5.05	.17	3.58	.17	3.58	.19	4.00
9	20	.10	2.11	.14	2.95	.22	4.63	.15	3.16	.12	2.53	.15	3.08
	50	.10	2.11	.15	3.16	.19	4.00	.15	3.16	.12	2.53	.14	2.99
10	20	.10	2.11	.25	5.27	.14	2.95	.14	2.95	.11	2.32	.15	3.12
	50	.14	2.95	.25	5.27	.18	3.79	.20	4.21	.20	4.21	.19	4.09
11	20	.10	2.11	.15	3.16	.10	2.11	.12	2.53	.13	2.74	.12	2.53
	50	.13	2.74	.16	3.37	.13	2.74	.17	3.58	.11	2.32	.14	2.95
12	20	.10	2.11	.13	2.74	.10	2.11	.34	7.16	.10	2.11	.15	3.25
	50	.10	2.11	.13	2.74	.10	2.11	.18	3.79	.20	4.21	.14	2.99

*ITBS/TAP

Table 2
Expected Growth Indices
in Standard Deviation Units & NCEs
Total Reading
Spring - Spring Testing

Grade	%ile	CAT '77		CTBS '75		ITBS '78*		MAT '78		SRA '78		AVG.	
		SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE
2	20	1.00	21.06	1.10	23.17	.60	12.64	1.20	25.27	1.00	21.06	.98	20.64
	50	1.20	25.27	1.30	27.38	1.07	22.53	1.20	25.27	.78	16.43	1.11	23.38
3	20	.70	14.74	.76	16.00	.81	17.06	.61	12.85	.63	13.27	.70	14.78
	50	.75	15.80	.88	18.50	.88	18.53	.53	11.16	.48	10.11	.70	14.82
4	20	.54	11.37	.48	10.11	.67	14.11	.45	9.48	.40	8.42	.51	10.70
	50	.70	14.74	.59	12.43	.72	15.16	.47	9.90	.60	12.64	.62	12.97
5	20	.42	8.85	.41	8.64	.58	12.21	.39	8.21	.17	3.58	.39	8.30
	50	.45	9.48	.48	10.11	.68	14.32	.31	6.53	.35	7.37	.45	9.56
6	20	.35	7.37	.21	4.42	.45	9.48	.10	2.11	.40	8.42	.30	6.36
	50	.40	8.42	.33	6.96	.55	11.58	.24	5.05	.60	12.64	.42	8.93
7	20	.30	6.32	.27	5.69	.43	9.06	.32	6.73	.38	8.00	.34	7.16
	50	.33	6.95	.33	6.95	.48	10.11	.20	4.21	.25	5.27	.32	6.70
8	20	.33	6.95	.31	6.53	.36	7.58	.30	6.32	.41	8.64	.34	7.20
	50	.39	8.21	.29	6.11	.44	9.27	.28	5.90	.27	5.69	.33	7.04
9	20	.19	4.00	.22	4.63	.28	5.90	.19	4.00	.25	5.27	.23	4.76
	50	.23	4.84	.26	5.48	.32	6.74	.20	4.21	.11	2.32	.22	4.72
10	20	.24	5.05	.37	7.79	.18	3.79	.29	6.11	.14	2.95	.24	5.14
	50	.30	6.32	.39	8.21	.25	5.27	.30	6.32	.30	6.32	.31	6.49
11	20	.24	5.05	.26	5.48	.17	3.58	.16	3.37	.35	7.37	.24	4.97
	50	.23	4.84	.24	5.05	.21	4.42	.23	4.84	.28	5.90	.24	5.01
12	20	.10	2.11	.23	4.84	.17	3.58	.36	7.58	.13	2.74	.20	4.17
	50	.14	2.95	.26	5.48	.17	3.58	.20	4.21	.31	6.53	.22	4.55

Table 3
Expected Growth Indices
in Standard Deviation Units & NCEs
Total Math
Fall - Spring Testing

Grade	%ile	CAT '77		CTBS '75		ITBS '78*		MAT '78		SRA '78		AVG.	
		SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE
2	20	.76	16.01	.56	11.79	.73	15.37	.51	10.74	.77	16.22	.67	14.08
	50	.83	17.47	.62	13.06	.73	15.37	.63	13.27	1.03	21.69	.77	16.17
3	20	.80	16.85	.52	10.95	.64	13.47	.29	6.11	.79	16.64	.61	12.80
	50	1.00	21.06	.65	13.69	.64	13.48	.45	9.48	.88	18.53	.72	15.25
4	20	.38	8.00	.14	2.95	.50	10.53	.50	10.53	.25	5.27	.35	7.46
	50	.53	11.16	.19	4.00	.58	12.22	.48	10.11	.38	8.00	.43	9.10
5	20	.33	6.95	.30	6.32	.54	11.37	.46	9.69	.31	6.53	.39	8.17
	50	.51	10.74	.36	7.58	.54	11.37	.40	8.42	.39	8.21	.44	9.26
6	20	.24	5.05	.14	2.95	.47	9.90	.14	2.95	.60	12.64	.32	6.70
	50	.40	8.42	.19	4.00	.47	9.90	.24	5.05	.41	8.63	.34	7.20
7	20	.17	3.58	.13	2.74	.41	8.63	.28	5.90	.17	3.58	.23	4.89
	50	.28	5.90	.17	3.58	.41	8.63	.26	5.48	.22	4.63	.27	5.64
8	20	.13	2.74	.11	2.32	.30	6.32	.09	1.90	.29	6.11	.18	3.88
	50	.21	4.42	.17	3.58	.30	6.32	.20	4.21	.32	6.74	.24	5.05
9	20	.10	2.11	.15	3.16	.18	3.79	.19	4.00	.10	2.11	.14	3.03
	50	.12	2.53	.13	2.74	.18	3.79	.14	2.95	.11	2.32	.14	2.87
10	20	0	0	.17	3.58	.13	2.74	.20	4.21	.10	2.11	.12	2.53
	50	.10	2.11	.17	3.58	.17	3.58	.13	2.74	0	0	.11	2.40
11	20	0	0	.10	2.11	.08	1.69	.10	2.11	.12	2.53	.08	1.69
	30	0	0	.08	1.69	.13	2.74	.10	2.11	.21	4.42	.10	2.19
12	20	-.10	-2.11	.05	1.05	.08	1.69	.10	2.11	.10	2.11	.05	.97
	50	0	0	.10	2.11	.08	1.69	.13	2.74	.10	2.11	.08	1.73

*ITBS/TAP

Table 4

Expected Growth Indices
in Standard Deviation Units & NCEs
Total Math
Spring - Spring Testing

Grade	%ile	CAT '77		CTBS '75		ITBS '78 *		MAT '78		SRA '78		AVG.	
		SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE	SD	NCE
2	20	1.30	27.38	1.10	23.17	1.30	27.38	1.00	21.06	.94	19.80	1.13	23.76
	50	1.40	29.48	1.50	31.59	1.30	27.38	1.10	23.17	1.31	27.59	1.32	27.84
3	20	1.10	23.17	1.10	23.17	1.30	27.38	.60	12.64	1.12	23.59	1.04	21.99
	50	1.30	27.38	1.30	27.38	1.30	27.38	.75	15.80	1.41	29.70	1.21	25.53
4	20	.70	14.74	.49	10.32	.92	19.38	.62	13.06	.40	8.42	.63	13.18
	50	.83	17.48	.55	11.58	1.10	23.17	.62	13.05	.60	12.64	.74	15.58
5	20	.62	13.06	.52	10.95	.92	19.38	.54	11.37	.43	9.06	.61	12.76
	50	.72	15.16	.68	14.32	.92	19.38	.46	9.69	.52	10.95	.66	13.90
6	20	.43	9.06	.30	6.32	.67	14.11	.20	4.21	.77	16.22	.47	9.98
	50	.53	11.16	.42	8.85	.80	16.85	.35	7.37	.60	12.64	.54	11.37
7	20	.25	5.27	.30	6.32	.59	12.43	.65	13.70	.24	5.05	.41	8.55
	50	.43	9.06	.37	7.79	.71	14.95	.39	8.21	.32	6.74	.44	9.35
8	20	.43	9.06	.23	4.84	.42	8.85	.18	3.79	.39	8.21	.33	6.95
	50	.50	10.53	.33	6.95	.53	11.16	.31	6.53	.44	9.27	.42	8.89
9	20	.24	5.05	.30	6.32	.28	5.90	.21	4.42	.20	4.21	.25	5.18
	50	.29	6.11	.33	6.95	.32	6.74	.18	3.79	.20	4.21	.26	5.56
10	20	.18	3.79	.29	6.11	.17	3.58	.21	4.42	.10	2.11	.19	4.00
	50	.25	5.27	.35	7.37	.26	5.48	.13	2.74	.11	2.32	.22	4.64
11	20	.15	3.16	.21	4.42	.17	3.58	.12	2.53	.19	4.00	.17	3.54
	50	.16	3.37	.16	3.37	.17	3.58	.13	2.74	.32	6.74	.19	3.96
12	20	0	0	.14	2.95	.12	2.53	.10	2.11	.11	2.32	.09	1.98
	50	.10	2.11	.18	3.79	.12	2.53	.16	3.37	.18	3.79	.15	3.12

*ITBS/TAP

Another contrast of interest in Figures 2-5 is that of an "average" (i.e., 50th percentile) student with a below average (i.e., 20th percentile) student. To the extent that these lines are not coincident, the initial gap between these students will widen or narrow over time. Most often the EG of a 50th percentile student exceeds that of the compensatory education student, particularly in math (see Figures 4 and 5). This results in a widening of the gap between them over time. The consistency of this trend is less in reading, particularly in the junior high years (grades 7-9), during which the 20th percentile student actually grows more to maintain that status than does the average student.

Insert Figures 2-5 here

Beyond interest in comparisons such as these, the information presented in Figures 2-5 has serious implications for the use of EG in obtaining a notion of percent additional growth (PAG), as discussed earlier. Clearly, a given NCE gain can reflect far greater or less achievement in relation to growth expectation depending upon the grade level (and pretest status) at which it is observed. For example, an NCE gain of 5 in a compensatory reading program (i.e., 20th percentile students) testing annually represents twice the percent additional growth in grade 9 (EG about 5) than it does in grade 5 (EG about 10, see Figure 3). Comparisons across a wider span of grades (or testing intervals, or subject matters) yield far more dramatic differences.

This type of interpretation--achievement gains in relation to expected growth--has been offered as a potentially useful means of communicating achievement gains to less technical audiences (Gabriel,

FIGURE 2

EXPECTED GROWTH

across five tests
in total READING

FALL - SPRING TESTING

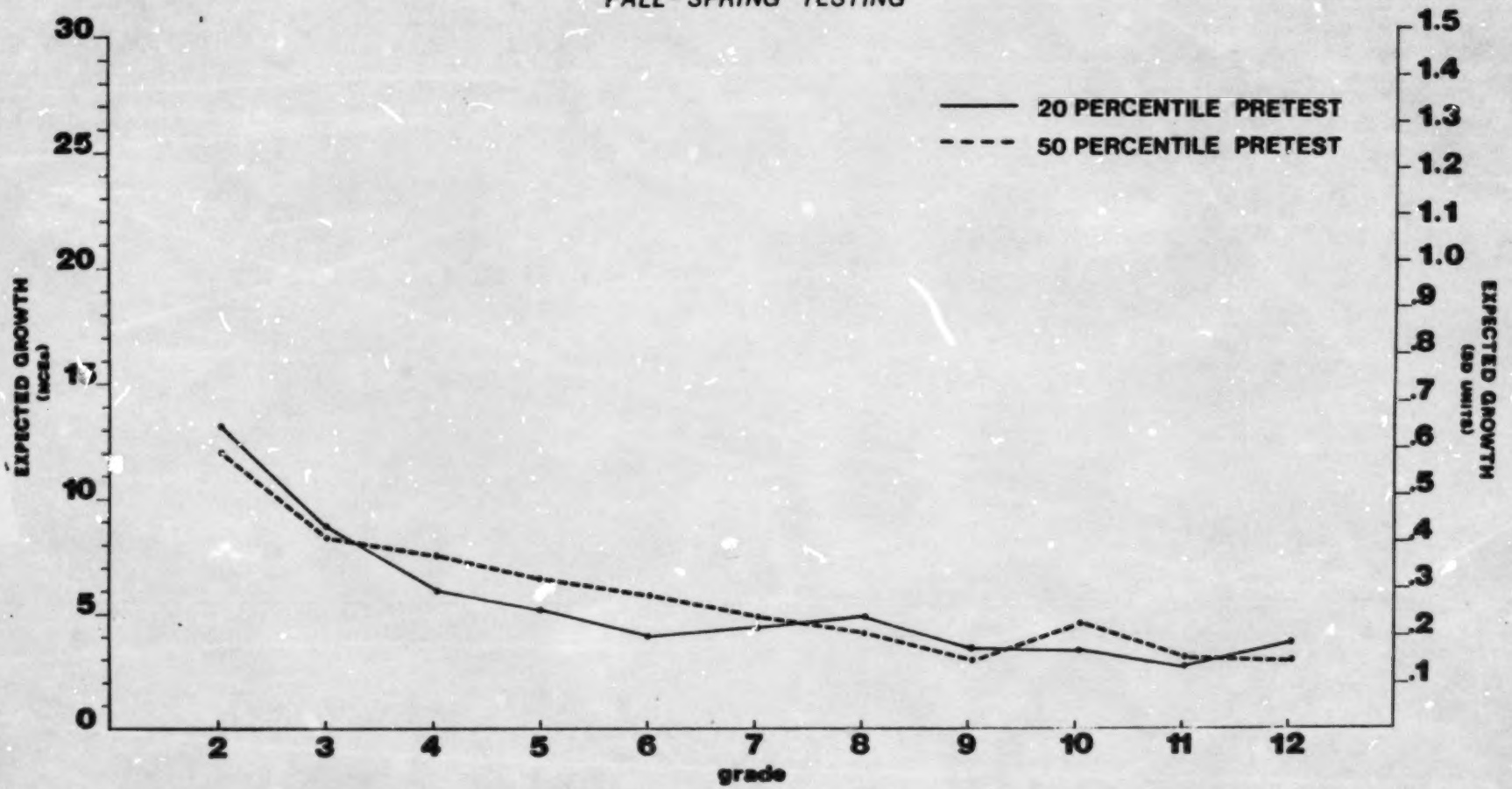


FIGURE 3

EXPECTED GROWTH

across five tests

in total READING

SPRING - SPRING TESTING

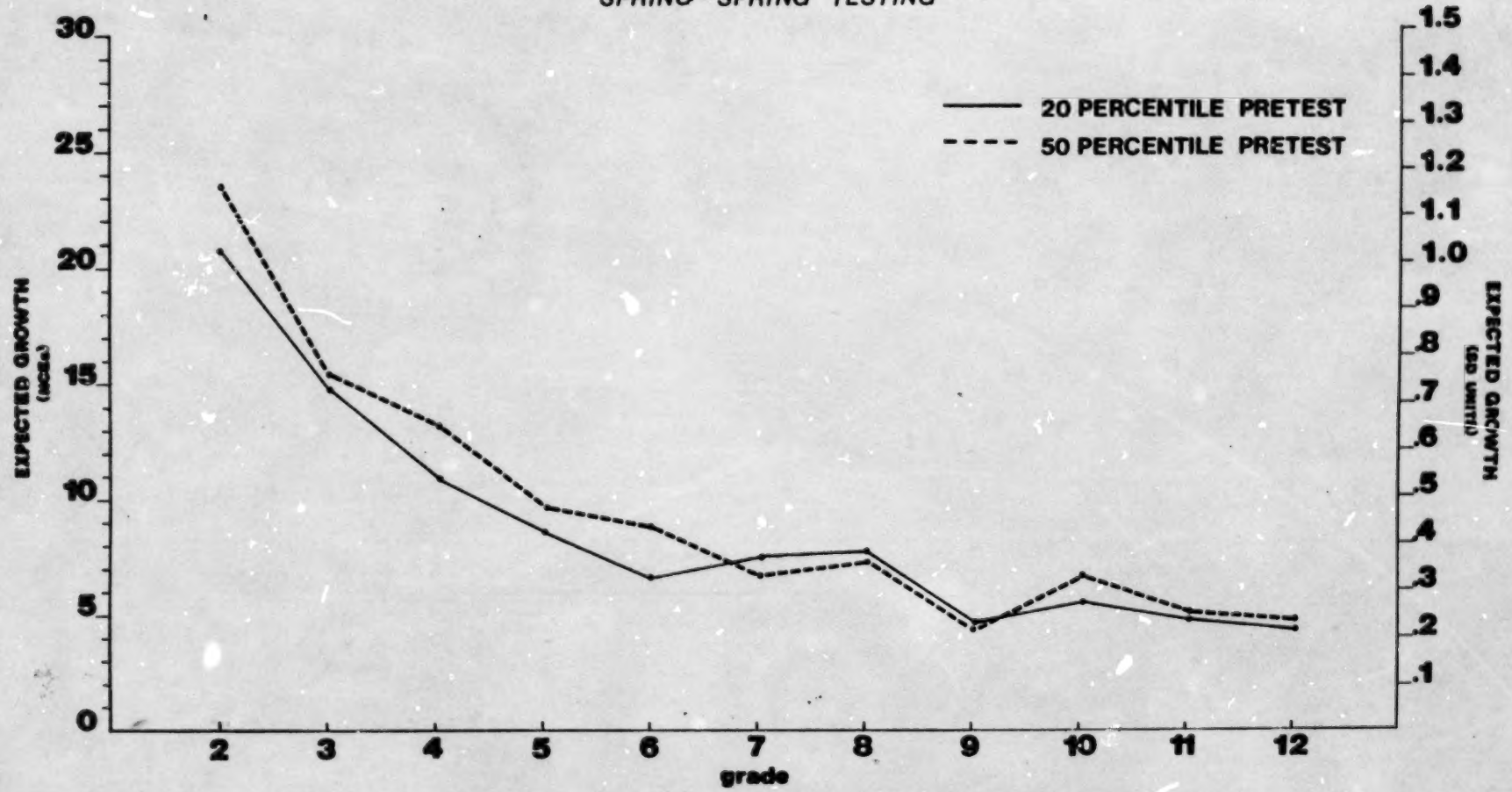
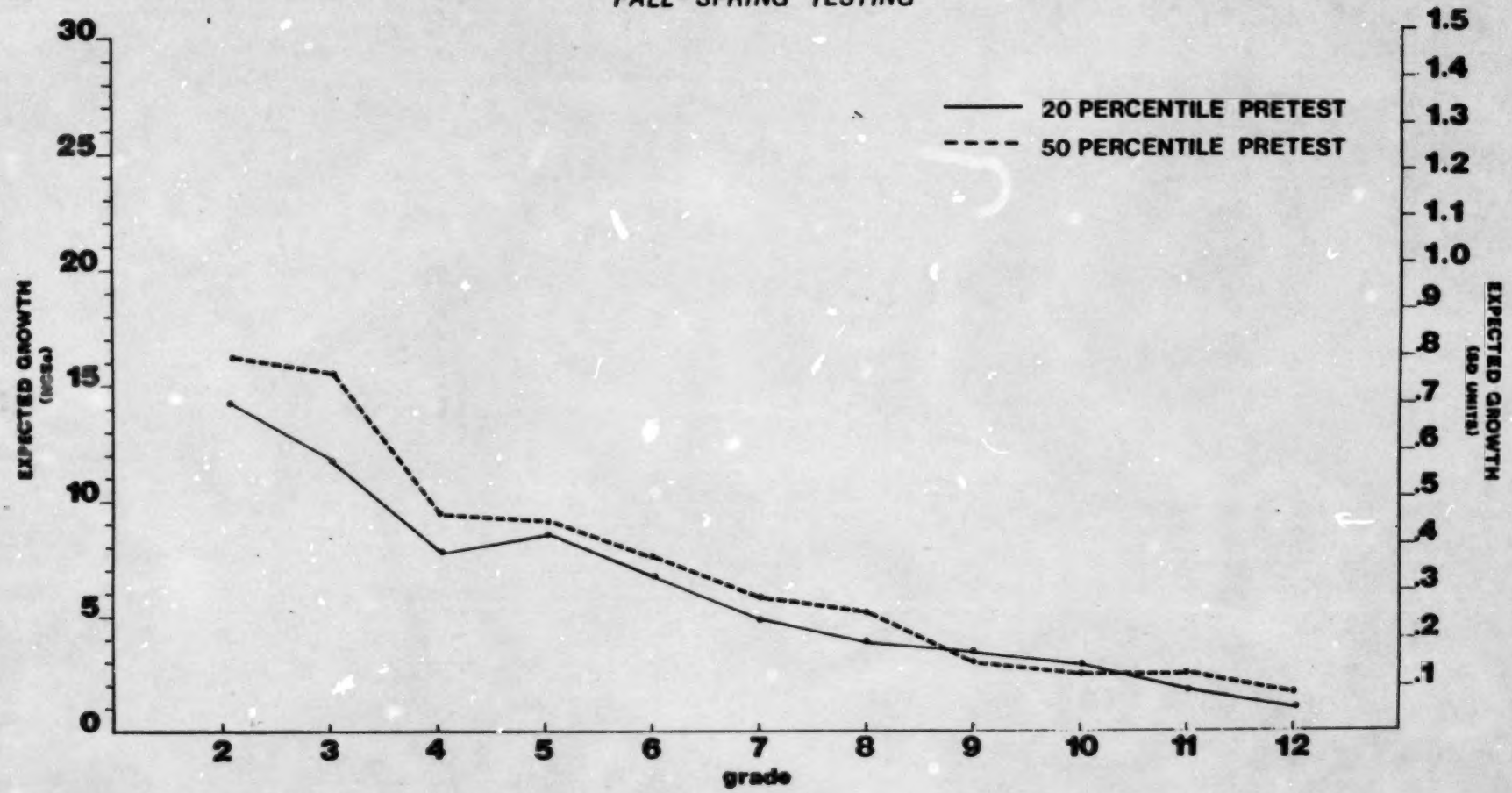


FIGURE 4

EXPECTED GROWTH

across five tests
in total MATH

FALL-SPRING TESTING



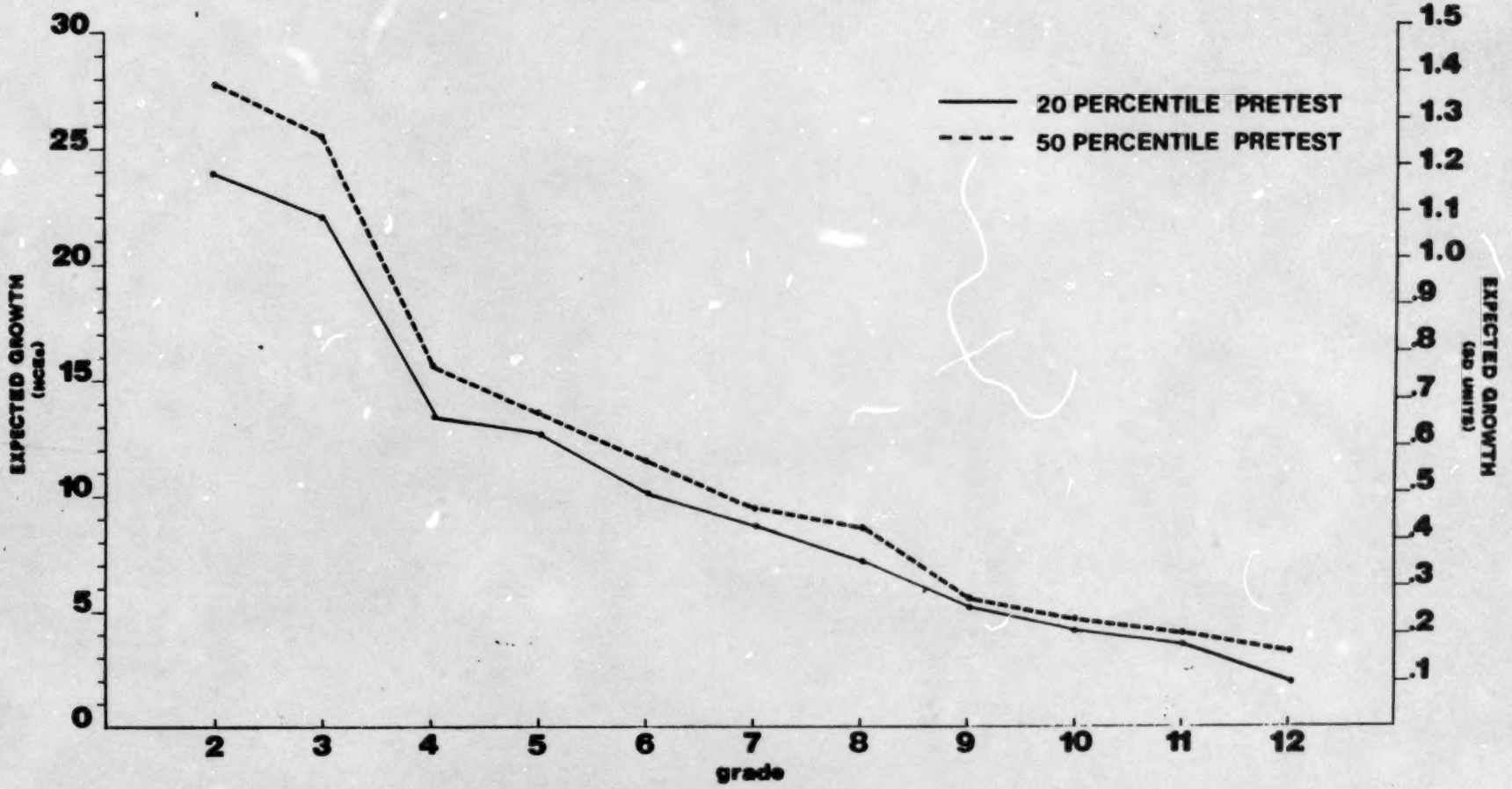
17

FIGURE 5

EXPECTED GROWTH

across five tests in total MATH

SPRING - SPRING TESTING



1981). It incorporates differences in growth rate across grade levels, and testing intervals for a given subject area. Recall, however, that the values represented in Figures 2-5 are average growth expectations for the five tests represented in Tables 1-4. To the extent that particular tests deviate from these average values, the growth expectations previously discussed are inaccurate for a given test or a different combination of tests. At the local level, where only one test is employed, this is obviously a crucial consideration. The solution, however, is simple--use the EG values for that particular test in communicating achievement gains. At a more aggregate level (e.g., SEA, federal), the situation is more complex. If tests differ dramatically in their growth expectations at a given grade, it becomes important to know how many tests, and in what relative frequency, are represented in the aggregate collection of project evaluation results

In the context of an aggregate evaluation reporting system, the issue of variation among tests in the interpretation of evaluation results is by no means a new one. The massive Anchor Test Study (Loret, Seder, Bianchini & Vale, 1972) employed a complex inter-test equating design to determine the equivalence of a given percentile ranking across eight reading comprehension tests in grades 4, 5 and 6. Jaeger (1977) used these data to illustrate the potential errors in interpreting the same score obtained on different tests as though it reflected the same level of performance, i.e., combining achievement results across tests in an aggregate reporting system such as TIERS. Yet, the need to devise a large-scale evaluation and reporting system general enough to provide information across a variety of particular characteristics of individual projects (e.g., hours per week of instruction, class size, test used) was

the fundamental charge in the development of TIERS (Wisler & Anderson, 1979). Tallmadge & Horst (1978) acknowledged the existence of differences among tests, such as those found in the Anchor Test Study, but placed the need for content validity of a particular test for a particular Title I program as paramount. That is, if each project employed the test which was most sensitive to its instructional objectives, equivalent achievement gains from different tests reflected equivalent progress through the curricula most relevant to those tests.

Roberts (1981) took a more psychometric approach to variations among tests such as those found in the Anchor Test Study. For the tests included in that study, he maintained that, while there were discrepancies in their equivalencies at individual percentile points, the differences between pairs of percentile points were virtually the same. That is, each tests' equivalency tables differed from the others by a constant amount across all percentiles. This provided support for maintaining that achievement gains, if not the pretest or posttest scores themselves, were equivalent across tests and could thus be aggregated. This, of course, provided crucial support to one of the intended products of the TIERS--the aggregation of achievement gains across Title I projects, LEAs and SEAs representing a variety of project characteristics.

On testing conditions more narrowly defined than those represented in the current paper, Gabriel (1981) noted serious differences in expected growth across the reading and math tests studied here, particularly for compensatory education students. Potential sources of these discrepancies included any of a number of normative characteristics of the tests and test levels themselves. The representativeness of the normative population, the precision of the norming process, the resultant

distribution of item difficulties, and the inter-level scaling properties were all potentially key factors in the variations across tests' growth expectations reported.

To address this issue in the five tests included in this study, the growth expectations for each test contained in Tables 1-4 are presented graphically in Figures 6-13. Perhaps a look at them for this sample of tests will illustrate the extent of the discrepancies among tests and subsequent interpretive pitfalls associated with PAG. Rather than discussing each of the eight figures individually, some general observations will be offered which the reader can confirm or refute by inspecting the graphs themselves. To better identify the information presented, consider Figure 6 as an example. It contains the EG values in a fall-spring testing interval for a 20th percentile student in reading. The ordinate expresses these values both in standard deviation units (far right) and in NCEs (far left). The solid line extending across grade levels in the body of the graph indicates the average growth expectation among the five tests at each grade. Each test's EG at each grade is plotted using the abbrevation of the test itself. For example, at grade 5 the MAT is at about the average of the five tests--an EG of approximately 6 NCEs. A ± 2 NCE band ($\pm .1$ standard deviation) around the average EG line is shown to provide some perspective in judging differences between the tests both within and across grade levels.

In Figures 6-13, variations among the five tests can be viewed in terms of the other factors represented: reading vs. math, fall-spring vs. spring-spring testing, 20th vs. 50th percentile students and

interactions among these conditions. Some general observations, by no means exhaustive, are offered here:

- The variation among tests is more pronounced in math than in reading.
- Annual testing evidences a wider range of EG values at a given grade than does fall-spring testing.
- Average students' EG is more variable than compensatory education students'.
- Variation among tests is most dramatic in the early grades.
- The convergence of EG values across tests comes at an earlier grade level in reading than in math.
- In reading, a given test's EG oscillates about the mean across grade levels. No test is consistently above or below the average.
- In math, tests more consistently have relatively high or low EG values across grades.

Insert Figures 6-13 about here

The differences among tests in all of these contrasts raise concerns for the generality of the growth expectation concept. One is tempted to conclude that, in interpreting achievement gains, a percent additional growth formulation is highly test and grade level specific. Again, this is less of a problem for local application, where these interpretations are made at the level of specificity anyway, than it is for a large scale, centralized reporting system like TIERS.

The range of growth expectations among tests at certain grade levels is striking. As mentioned above, these are most extreme in math at the early grade levels in an annual testing schedule (see Figures 11 and 13). Ranges of 10-15 NCEs in EG are not uncommon here--and this represents half a standard deviation and more! Thus, local use of PAG as

FIGURE 6

EXPECTED GROWTH

grades 2–12

total READING, 20th percentile

FALL-SPRING TESTING

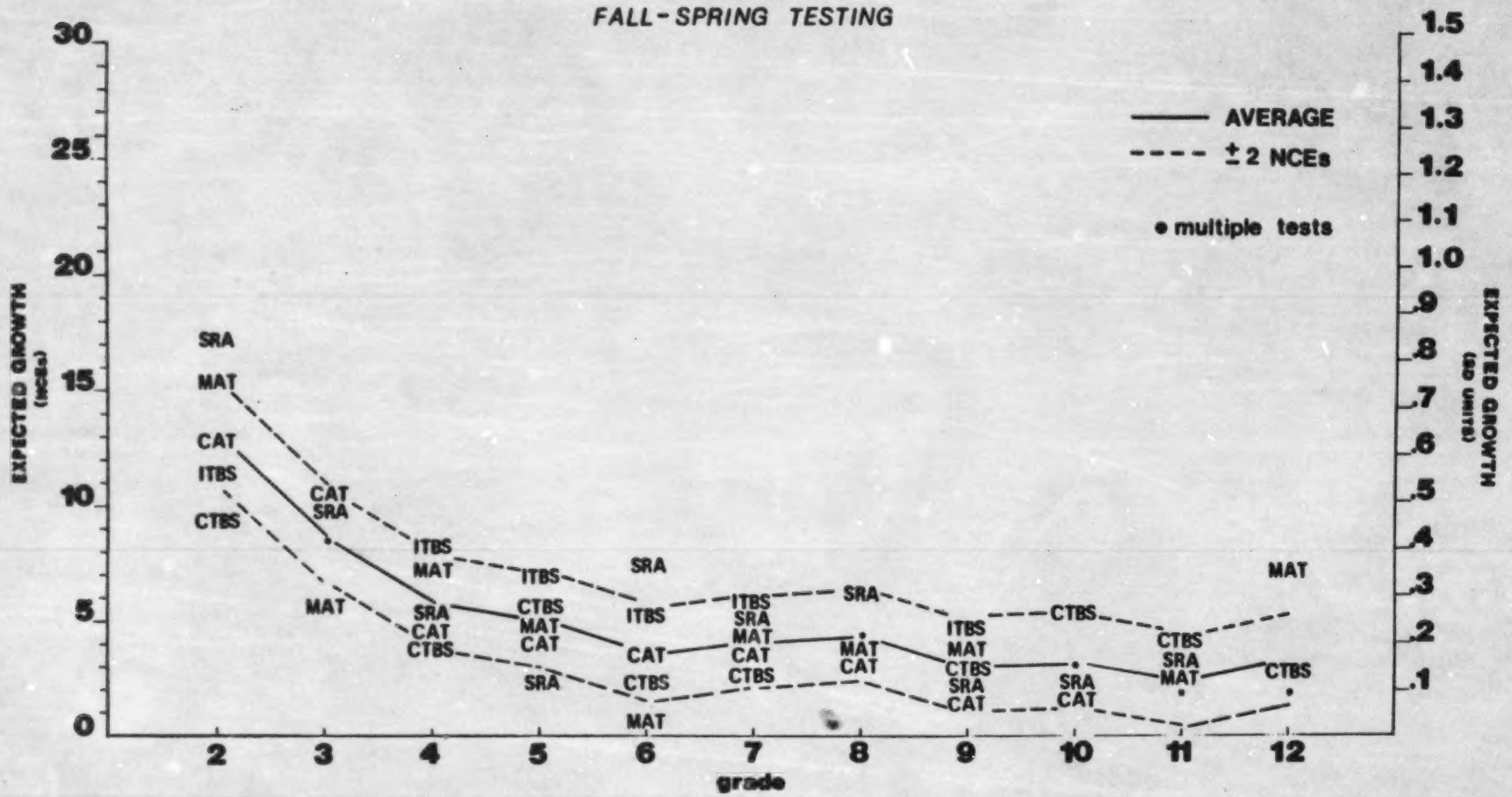


FIGURE 7

EXPECTED GROWTH

grades 2–12

total READING, 50th percentile

FALL-SPRING TESTING

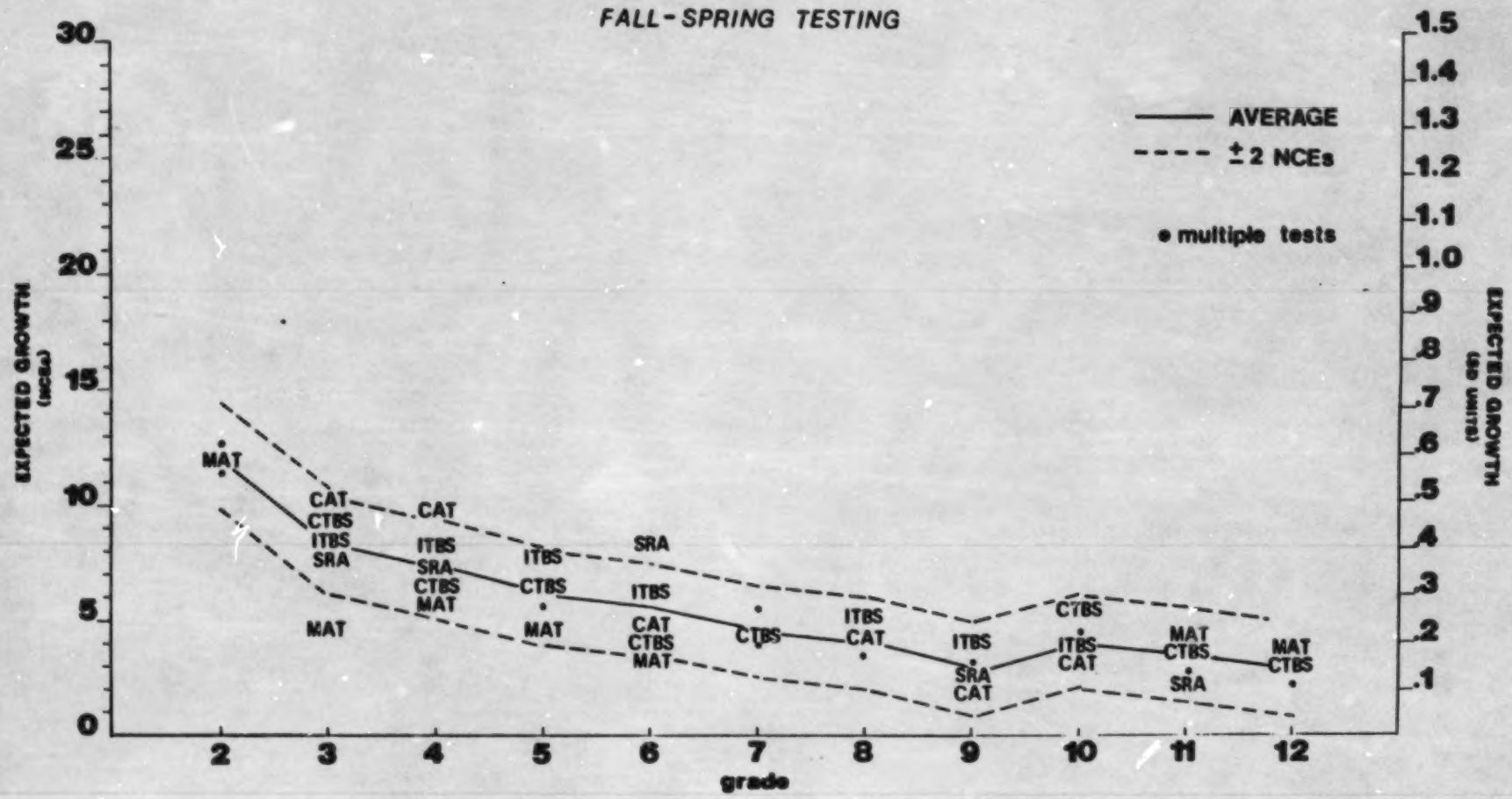


FIGURE 8

EXPECTED GROWTH

grades 2–12

total READING, 20th percentile

SPRING - SPRING TESTING

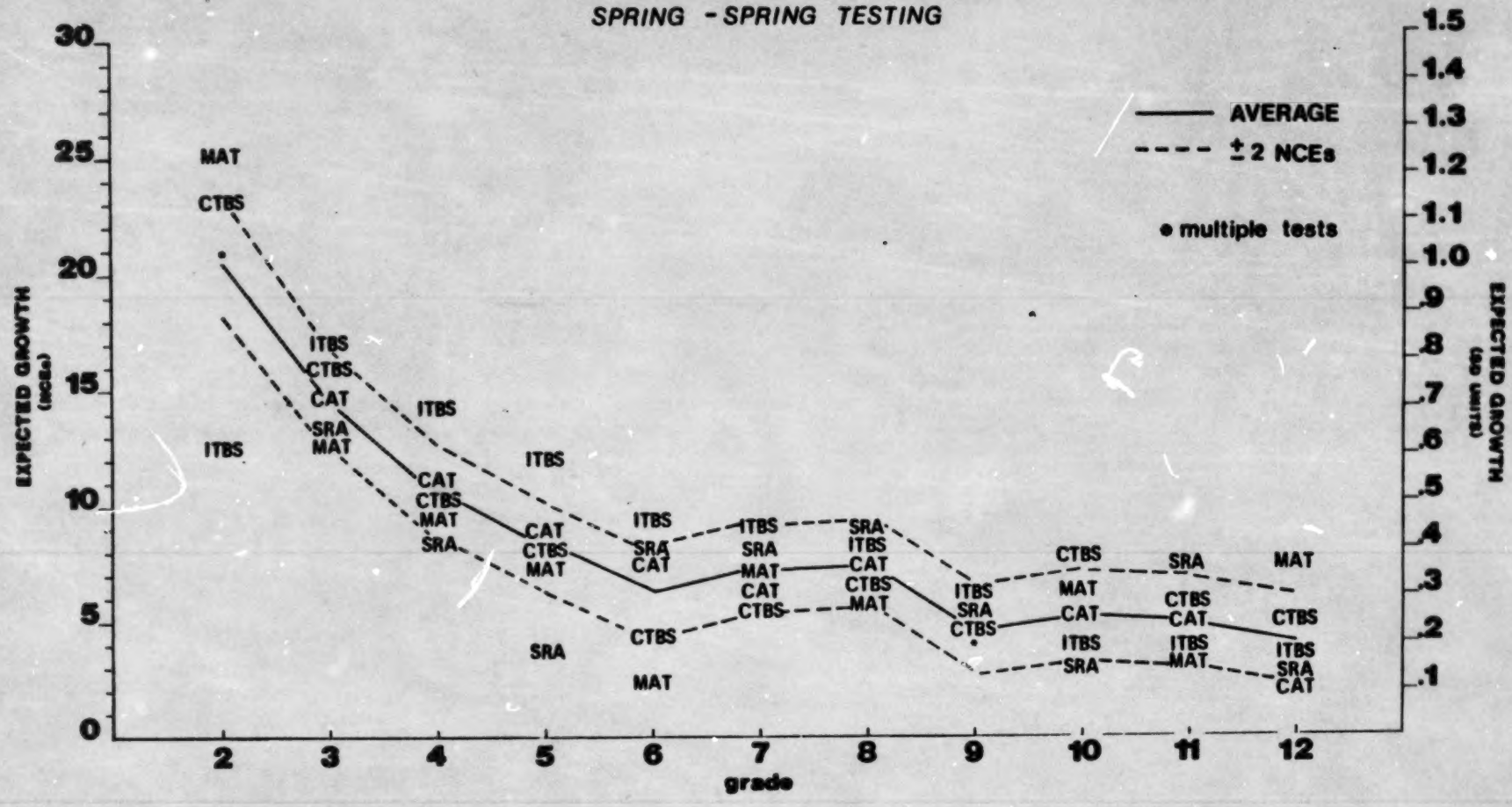


FIGURE 9

EXPECTED GROWTH

grades 2–12

total READING, 50th percentile

SPRING - SPRING TESTING

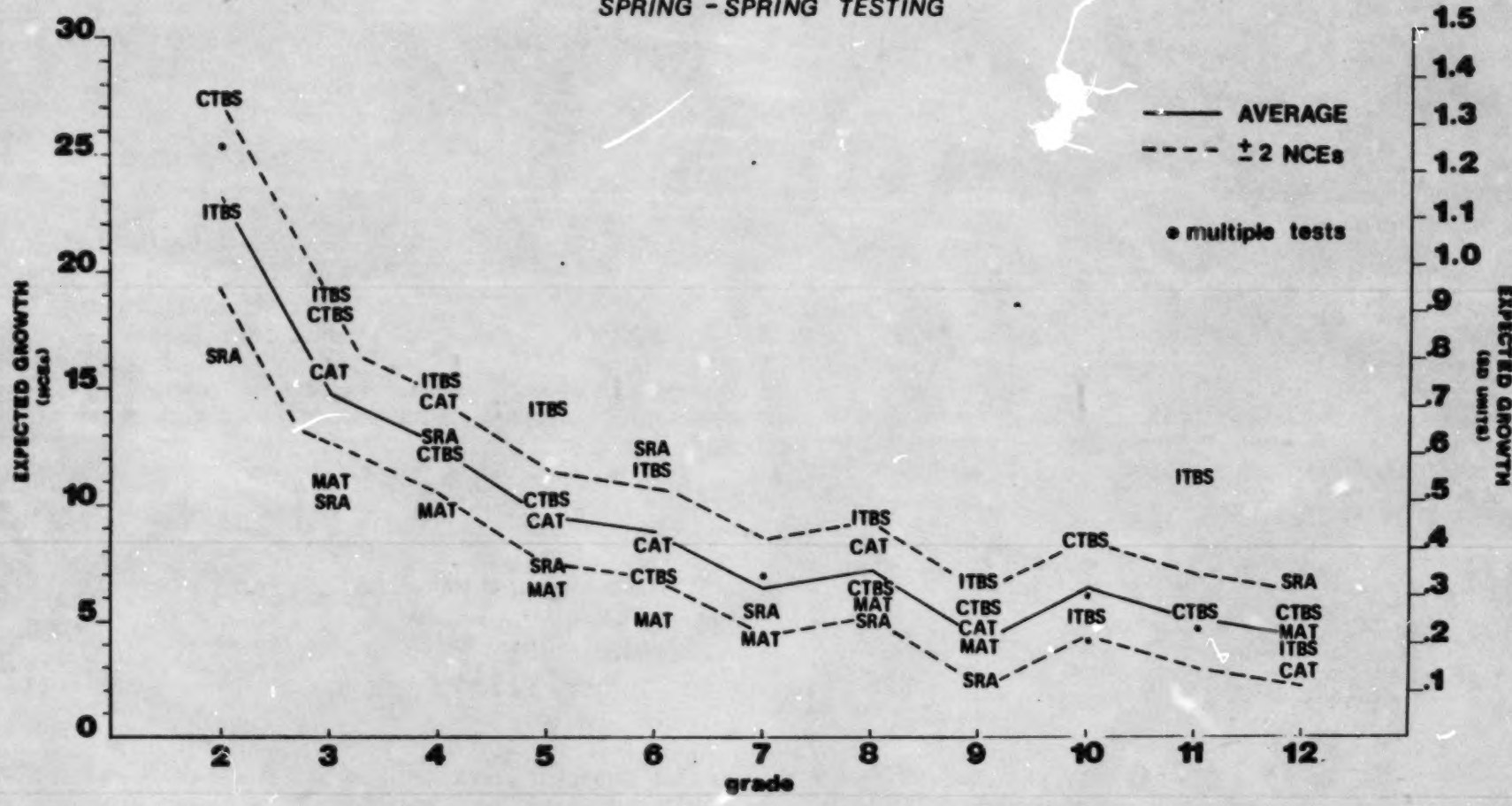


FIGURE 10

EXPECTED GROWTH

grades 2–12
in total MATH, 20th percentile

FALL-SPRING TESTING

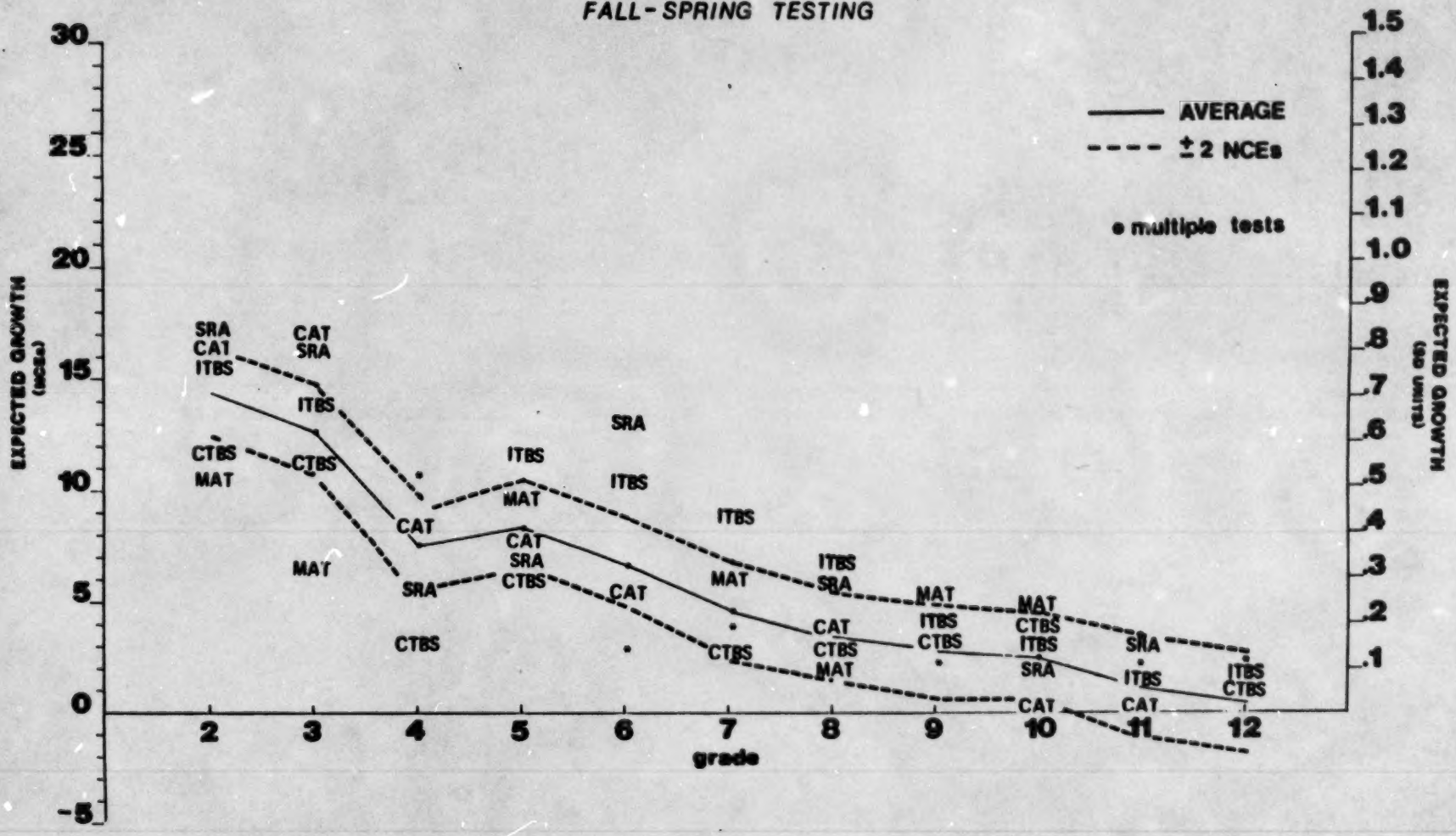


FIGURE 11

EXPECTED GROWTH

grades 2–12

in total MATH, .20th percentile

SPRING - SPRING TESTING

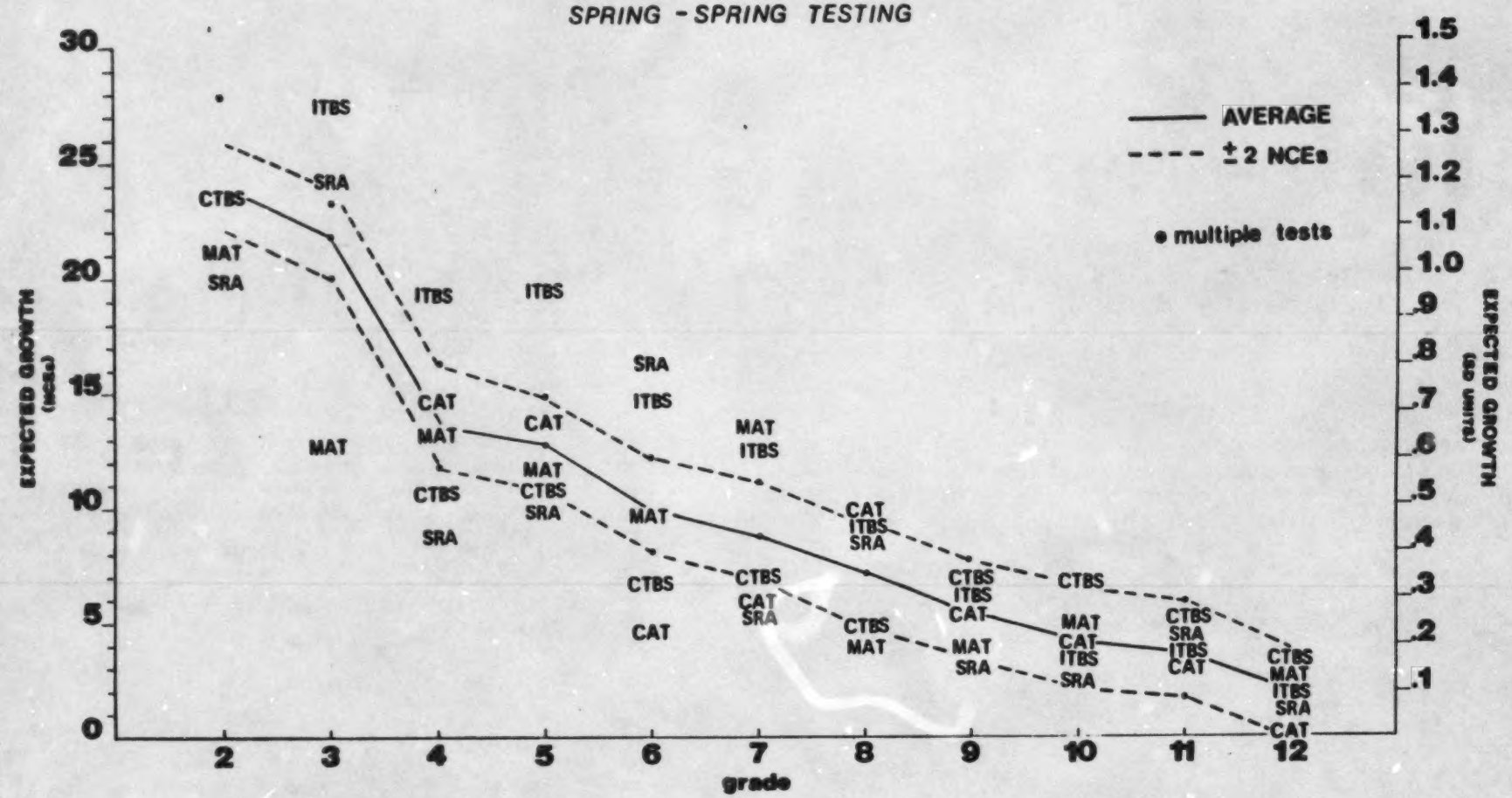


FIGURE 12

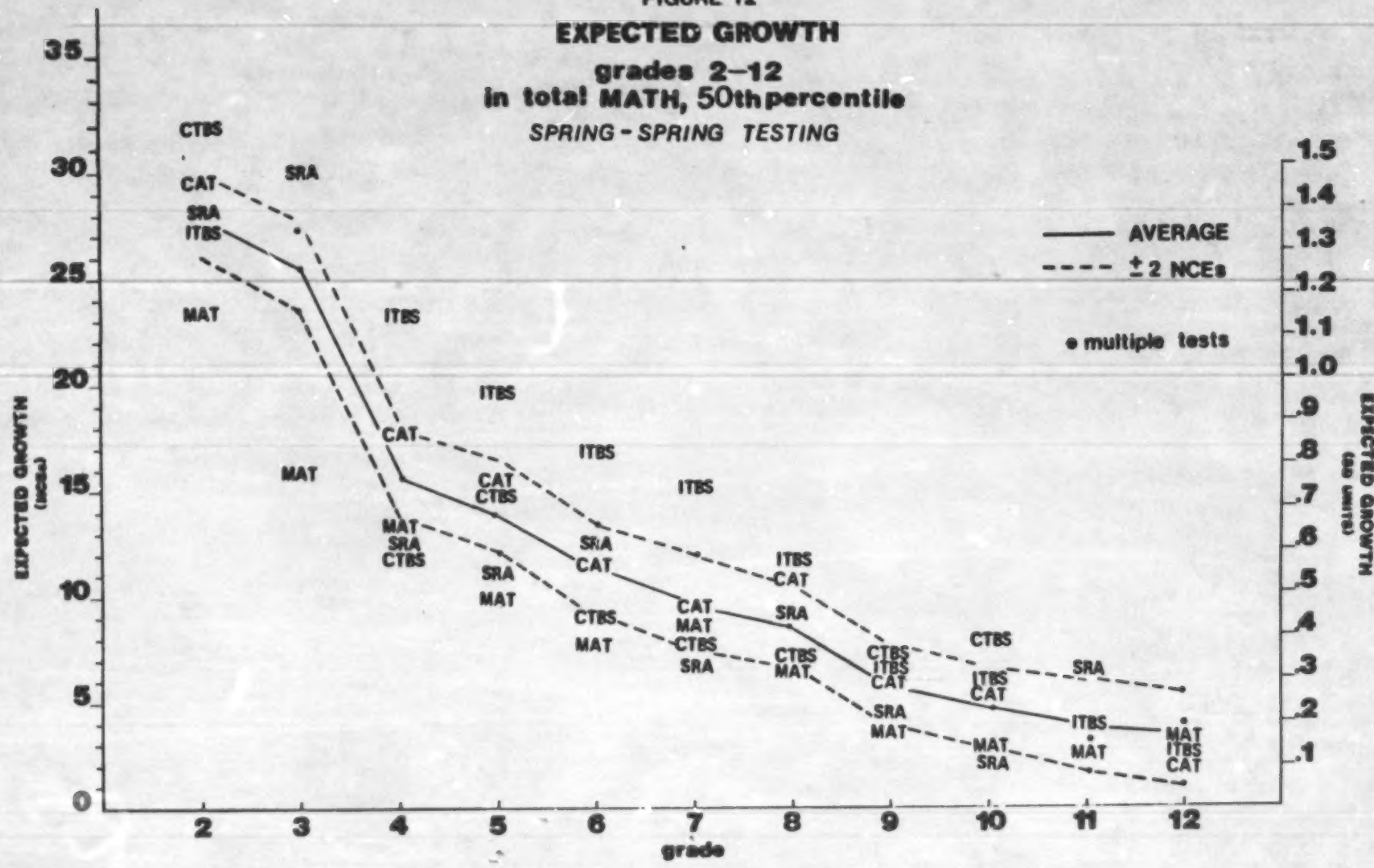
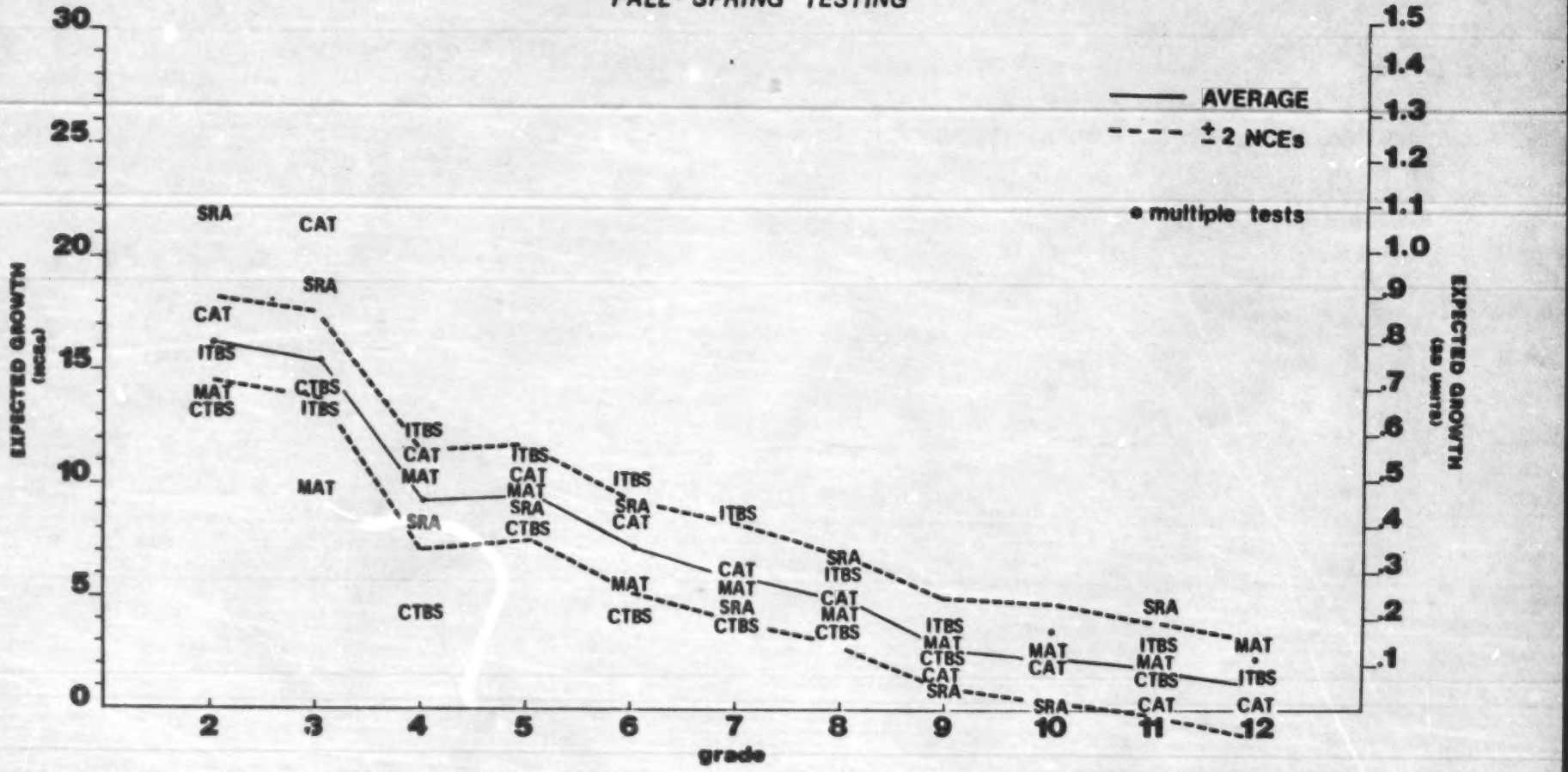

EXPECTED GROWTH
grades 2–12
in total MATH, 50th percentile
SPRING - SPRING TESTING
AVERAGE
± 2 NCEs
• multiple tests
EXPECTED GROWTH (NCEs)
EXPECTED GROWTH (GE UNITS)
grade
0
5
10
15
20
25
30
35
1.0
1.1
1.2
1.3
1.4
1.5
2
3
4
5
6
7
8
9
10
11
12
CTBS
CAT
SRA
ITBS
MAT

FIGURE 13

EXPECTED GROWTH

grades 2–12
in total MATH, 50th percentile

FALL-SPRING TESTING



an interpretive device is best advised to calculate EG specific to its own testing conditions. From a larger scale point of view, however, these extremes can often be misleading in attempting to judge the overall degree of generality/idiosyncrasy in the system.

To investigate the generality of growth expectations across the conditions represented here, a variety of regression analyses were performed. The extent to which the EG values in Tables 1-4 were predictable from knowledge of the other factors (grade level, pretest status, test used, etc.) was the focus of these analyses. Table 5 contains the simple correlations between EG and the predictor sets for reading and math. Included as predictors are the inverse of grade level to pick up the curvilinearity of the EG/grade relationship evident in Figures 6-13, and dummy variables representing a test x grade interaction to see if the EG/test relationship was moderated by the particular grade level in question.

Insert Table 5 about here

The magnitude of the overall relationship between grade level and growth expectation is an obviously striking result in Table 5. The relationship is more linear in math than in reading, as the reciprocal of grade adds more to predictability in the latter. Still, to attain the level of EG predictability represented by a correlation of .80 by simply specifying the grade level of interest appears to be quite a testament to the generality of the phenomenon across the tests, pretest levels, etc., represented here.

Table 5

Simple Correlations Between
Expected Growth and Predictor Set*
By Subject Matter

	Reading	Math
Grade	-.70	-.79
1/Grade	.80	.80
Pretest Status	-.07	-.09
Test Interval	-.37	-.31
Test (1)	.01	.01
Test (2)	.02	-.05
Test (3)	.06	.09
Test (4)	-.02	-.08
Test x Grade (1)	-.02	-.01
Test x Grade (2)	.01	-.03
Test x Grade (3)	.03	.07
Test x Grade (4)	-.02	-.04

*Based upon 220 EG values in each subject matter

By computing these correlations on data pooled across many of the conditions represented, there is a risk of artifically inflating (or sometimes deflating) the relationship of interest--much like pooling heterogeneous classrooms of students and calculating correlations at the student level (see, for example, Cronbach, 1976). The correlations in Table 5 were thus calculated separately for fall-spring and spring-spring testing intervals in each subject matter, and separately for 20th and 50th percentile status. In all cases, the relationships between EG and grade level grew stronger--approaching correlations of .90 with $^{1}/_{Grade}$ as predictor.

Insert Table 6 about here

In contrast to the strength of the EG/grade relationship are the relatively weak relationships between test and test x grade interaction predictors and growth expectation shown in Table 5. They are slightly higher in math, but still not as dramatic as one might predict from inspecting Figures 6-13. As with grade level, their correlations with EG increase when computed for pretest status and testing interval conditions separately--particularly the latter. These correlations are given in Table 6.

A variety of multiple regression analyses were performed on the data as well. Obviously, as long as grade level (or $^{1}/_{Grade}$) was in the equation, little else would be seen as having much predictive power. As indicated in Table 6, the reciprocal of grade level accounts for approximately 80% of the variance in growth expectation in reading and math for either testing interval. To close discussion of the influence

Table 6

Simple Correlations Between
Expected Growth and Predictor Set*
By Subject Matter and Testing Interval

	Reading		Math	
	F/S	S/S	F/S	S/S
Grade	-.77	-.79	-.84	-.87
1/Grade	.88	.91	.84	.89
Pretest Status	-.06	-.08	-.09	-.10
Test (1)	-.07	.05	-.04	.04
Test (2)	-.07	.07	-.15	.01
Test (3)	.01	.10	.04	.13
Test (4)	-.07	.00	-.10	-.07
Test x Grade (1)	-.07	.01	-.05	.01
Test x Grade (2)	-.03	.03	-.08	.01
Test x Grade (3)	.01	.05	.05	.09
Test x Grade (4)	-.02	-.02	-.04	-.04

*Based upon 110 EG values in each subject matter/test interval

of interest variation on EG, however, the relationship between test used and a test by grade interaction with growth expectation was investigated via multiple linear regression analyses. Table 7 contains results of these analyses, with multiple correlations of the four test dummy variables and EG, the four interaction dummy variables and EG, and both combined and EG.

Insert Table 7 about here

As noted in the simple correlations presented earlier, these relationships are stronger in math than in reading. As a package, the test and test x grade variables account for less than 5% of the variance in reading growth expectations and between 5% and 10% of the variance of math growth expectations. Some of this surprisingly weak relationship relates to an observation offered earlier in Figures 6-13. The tendency for a given test to oscillate about the mean EG through the grade levels necessarily precludes any consistent, linear relationship between test and EG across those grade levels. In reading (see Figures 6-9), it appears that there is as much variation <u>within</u> a test across the grade levels as there is <u>between</u> tests.

Having displayed and discussed all of the differences and contrasts in EG across the conditions represented in this study, the author would like to return to some notion of generality in dealing with academic growth expectations. Recalling the strength of the simple relationship between grade level (inverse) and EG, one might establish generalized growth expectations by grade level--and testing interval--on this basis

Table 7

Multiple Correlations Between Expected Growth
And Test and Test x Grade Interaction Variables

	Reading		Math	
	F/S	S/S	F/S	S/S
Test (1-4) only	.11	.16	.21	.21
Test x Grade (1-4) only	.09	.08	.14	.13
Test (1-4) and Test x Grade (1-4)	.20	.17	.30	.25

alone. The regression equations and standard errors of estimate are shown below:

Reading	(F/S)	$EG = .92 + \frac{23.32}{Grade}$	($s_e = 1.25$)
	(S/S)	$EG = .91 + \frac{42.69}{Grade}$	($s_e = 2.33$)
Math	(F/S)	$EG = -.18 + \frac{35.35}{Grade}$	($s_e = 2.74$)
	(S/S)	$EG = -.49 + \frac{60.37}{Grade}$	($s_e = 3.71$)

By simply substituting the grade level of interest in the equation, a growth expectation is derived for that grade. The results of these calculations, rounded to a whole NCE, are presented in Table 8. Admittedly, the data on which the equation was based is by no means exhaustive of the volumes of current test norms that exist. Yet the patterns in Table 8 have some face validity. The decline in EG with increasing grades is decelerating and gradually levels off. The discrepancy between fall-spring and annual growth expectations is at a roughly constant proportion across the grades. Fall-spring EG is approximately 60% of annual EG, and this corresponds closely with the time differential between the two testing intervals--fall-spring empirical norm dates are typically six months apart, while the school year is considered to be 10 months apart. If one were to continue to

strive for simplicity, this time differential could be built into the equation, necessitating only one equation for each subject matter.

Insert Table 8 about here

These rough estimates may be most useful at a central level, one receiving a large number of evaluation reports representing different testing conditions. Along with the need to produce an aggregate picture of program effectiveness (e.g., NCE gains across projects), the growth expectations shown in Table 8 may be of some interpretive value. A state coordinator may profit from the understanding that in math, third graders typically exhibit about twice as much annual growth (EG = 20, see Table 8) than sixth graders (EG = 10) with just regular schooling. Perhaps the oft-repeated pattern of high NCE gains in the earlier grades will become more understandable as consistent with normal growth patterns, and less an indictment of junior and senior high school compensatory education programs.

Table 8

Regression Estimates of NCE Growth Expectations

Grade	Reading		Math	
	F/S	S/S	F/S	S/S
2	13	22	17	29
3	8	15	12	20
4	7	12	8	15
5	6	9	7	12
6	5	8	6	10
7	4	7	5	8
8	4	6	4	7
9	4	6	4	6
10	3	5	3	6
11	3	5	3	5
12	3	4	2	4

REFERENCES

Cronbach, L.J. Research on classrooms and schools: Formulation of questions, designs and analysis. Stanford Evaluation Consortium Occasional Paper Series. Stanford, California, July, 1976.

Davis, A., Deck, D., & Demaline, R. Interpretation guide for Title I evaluation results. Northwest Regional Educational Laboratory ESEA Title I Technical Assistance Center. Portland, Oregon, 1981.

Estes, G.D. Estimating gains and establishing program objectives using data from nationally normed tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 1979.

Gabriel, R.M. Norm-referenced growth expectations: Interpretive cautions for a potentially useful index. Paper presented at the annual meeting of the California Educational Research Assocation, San Diego, CA, November, 1981.

Jaeger, R.M. On combining achievement test data through NCE scaled scores. Research Triangle Park, NC: Research Triangle Institute, 1977, 94.

Jaeger, R.M. The effect of test selection on Title I project impact. Educational Evaluation and Policy Analysis, 1979, 1, 33-40.

Linn, R.L. Validity of inferences based on the proposed Title I evaluation models. Educational Evaluation and Policy Analysis, 1979, 1, 23-32.

Loret, P.G., Seder, A., Bianchini, J.C., & Vale, C.A. Anchor Test Study: Final Report. Berkeley, California: Educational Testing Service, 1972.

Roberts, A.O.H. Comparison of gains across publishers and revisions. Mountain View, CA: RMC Research Corporation, March, 1981.

Stenner, A.J., Hunter, E.L., Bland, J.D., & Cooper, M.L. The standardized growth expectation: implications for educational evaluation. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada, 1978.

Tallmadge, G.K. Interpreting NCEs. ESEA Title I Evaluation and Reporting System Technical Paper No. 2. Mountain View, CA: RMC Research Corporation, October, 1978.

Tallmadge, G.K. & Fagan, B.M. Cognitive growth and growth expectations in reading and mathematics. Mountain View, CA: RMC Research Corporation, November, 1977.

Tallmadge, G.K. & Horst, D.P. The use of different achievement tests in the ESEA Title I evaluation system. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, 1978.

Tallmadge, G.K. & Wood, C.T. User's Guide. Mountain View, CA: RMC Research Corporation, October, 1976. (User's Guide to the ESEA Title I Evaluation and Reporting System)

Wiley, D.E. Evaluation by aggregation: social and methodological biases. Educational Evaluation and Policy Analysis, 1979, 1, 41-45.

Wisler, C.E. & Anderson, J.K. Designing a Title I evaluation system to meet legislative requirements. Educational Evaluation and Policy Analysis, 1979, 1, 47-55.